Analyzing Change Over Time

A visual diagnostic guide to the one-way repeated measures ANOVA.

Movement Science Statistical Diagnostics

NotebookLM

The Efficiency of Serving as Your Own Control

Between-Subjects Design

Each participant receives only one treatment.

The error term absorbs all individual differences—genetics, training history, and lifestyle.

This requires massive sample sizes (e.g., n=150) to detect medium effects.

Within-Subjects Design

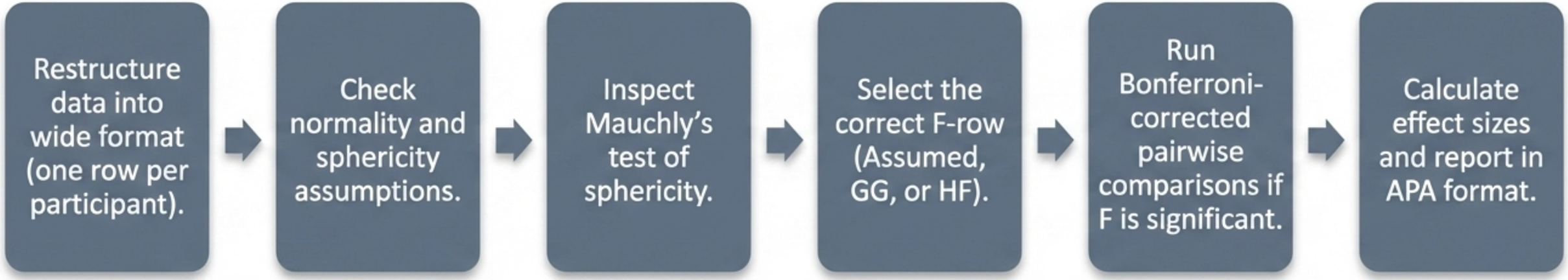
The same participants appear in all conditions.

The same participants appear in all conditions. Consistent baseline individual differences completely cancel out of the error term.

This achieves the same statistical power with a fraction of the participants (e.g., n=30).

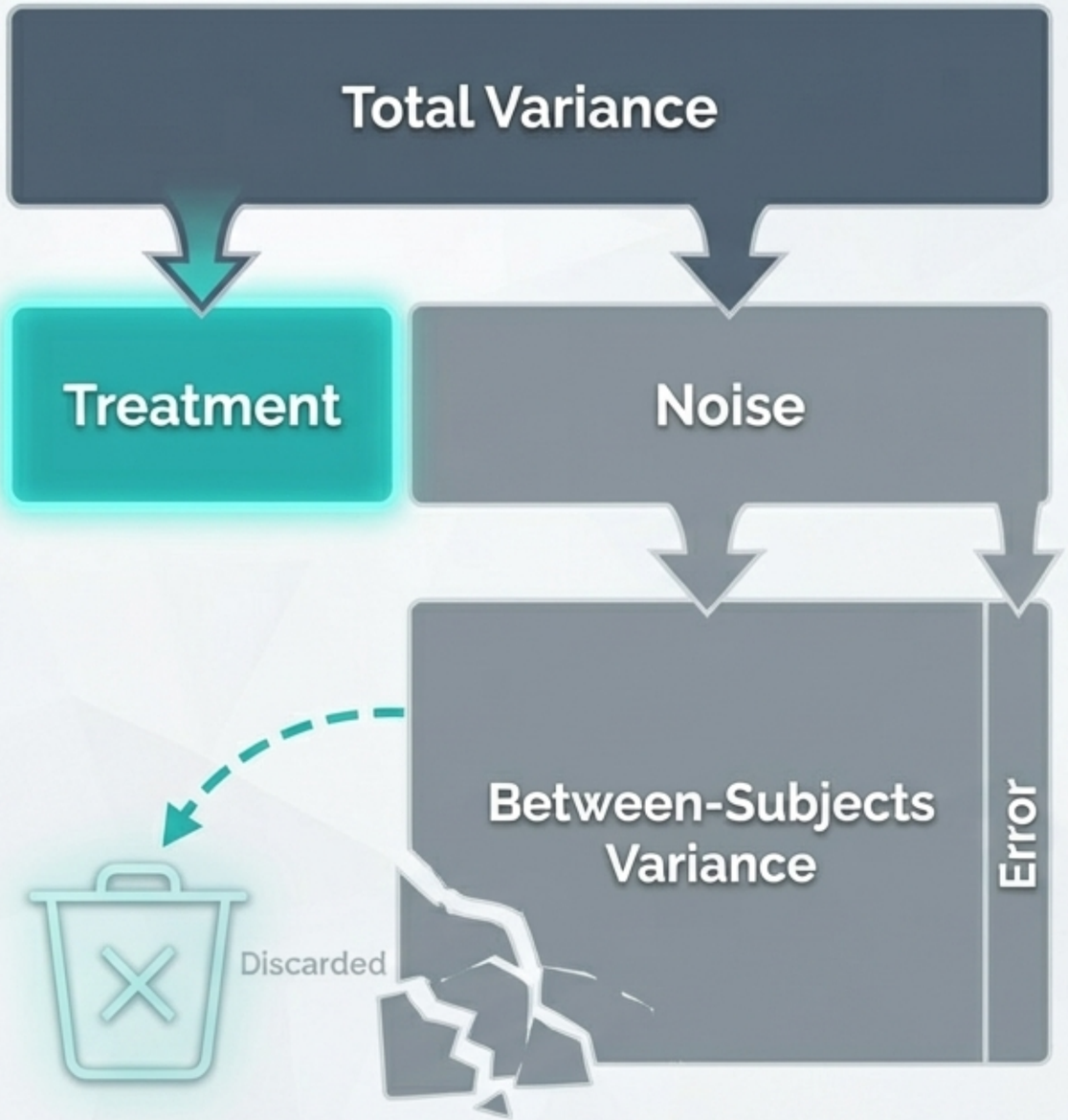
NotebookLM

The Analytical Pipeline


Restructure data into wide format (one row per participant).
Check normality and sphericity assumptions.
Inspect Mauchly's test of sphericity.
Select the correct F-row (Assumed, GG, or HF).
Run Bonferroni-corrected pairwise comparisons if F is significant.
Calculate effect sizes and report in APA format.

NotebookLM

The **Power Advantage:** Partitioning Out Individual Differences



The Theory

A smaller denominator produces a larger F-statistic.

$$F = \frac{\text{Treatment}}{\text{Error}}$$

By removing consistent individual baseline differences from the error term, we isolate genuine change.

Error

The Application

Participant A starts a strength program at 110 kg.
Participant B starts at 60 kg.

Rather than treating this massive 50 kg baseline difference as error, the analysis isolates only how much each person changed relative to their own starting point.

NotebookLM

The Case Study: A 12-Week Resistance Training Program

The null hypothesis assumes the population mean is equal at all time points:

$$\mu_{pre} = \mu_{mid} = \mu_{post}.$$

Pre-training

n = 30
Mean = 79.7 kg

Mid-training (6-week)

n = 30
Mean = 81.7 kg

Post-training (12-week)

n = 30
Mean = 85.1 kg

NotebookLM

Baseline Diagnostics: Independence and Normality

Independence

Different participants' scores must be independent of one another.

Note that scores within the same participant are naturally correlated—that is the explicit purpose of this design.

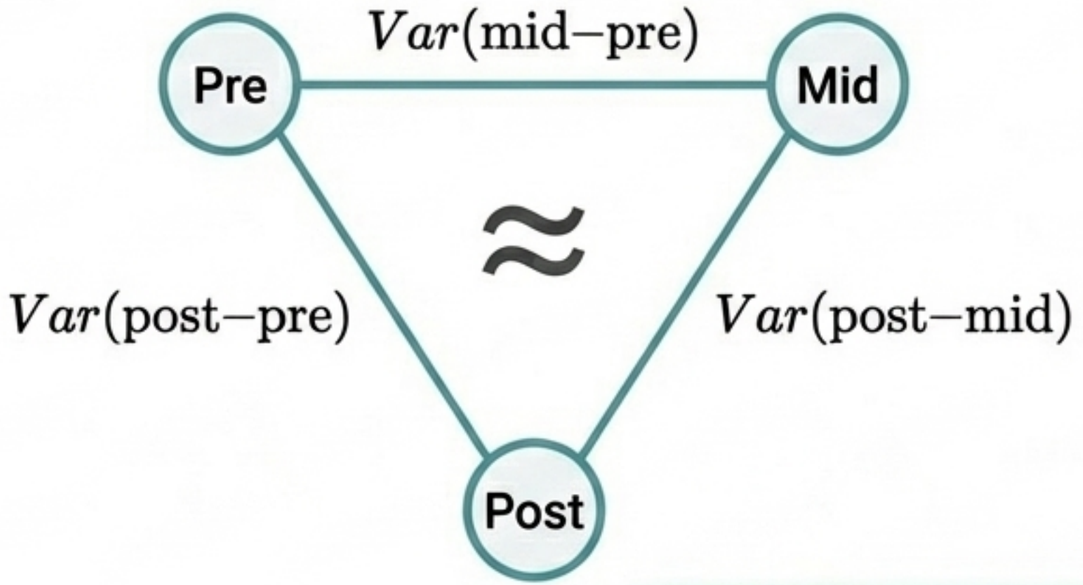
Normality of Difference Scores

Unlike between-subjects ANOVA, which requires normality of raw scores, repeated measures requires that the differences between each pair of time points are normally distributed.

You must verify that the distributions of (mid-pre), (post-pre), and (post-mid) are approximately normal.

NotebookLM

The Core Assumption: Sphericity



The Theory

Sphericity requires that the variances of the differences between all pairs of time points are approximately equal. The consistency of change must be similar across all time intervals.

Diagnostic Warning

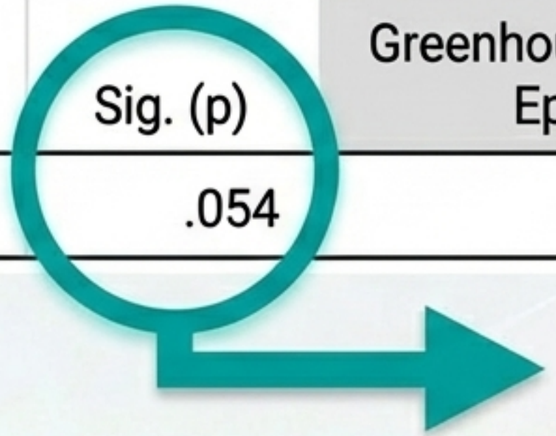
Sphericity is about the differences, not the raw scores. A dataset can have drastically different standard deviations at pre, mid, and post and still meet the assumption if the individual changes are uniform.

NotebookLM

Diagnosing Sphericity with Mauchly's W

Mauchly's Test of Sphericity

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig. (p)	Greenhouse-Geisser Epsilon
Time	.932	5.678	2	.054	.941



Interpretation

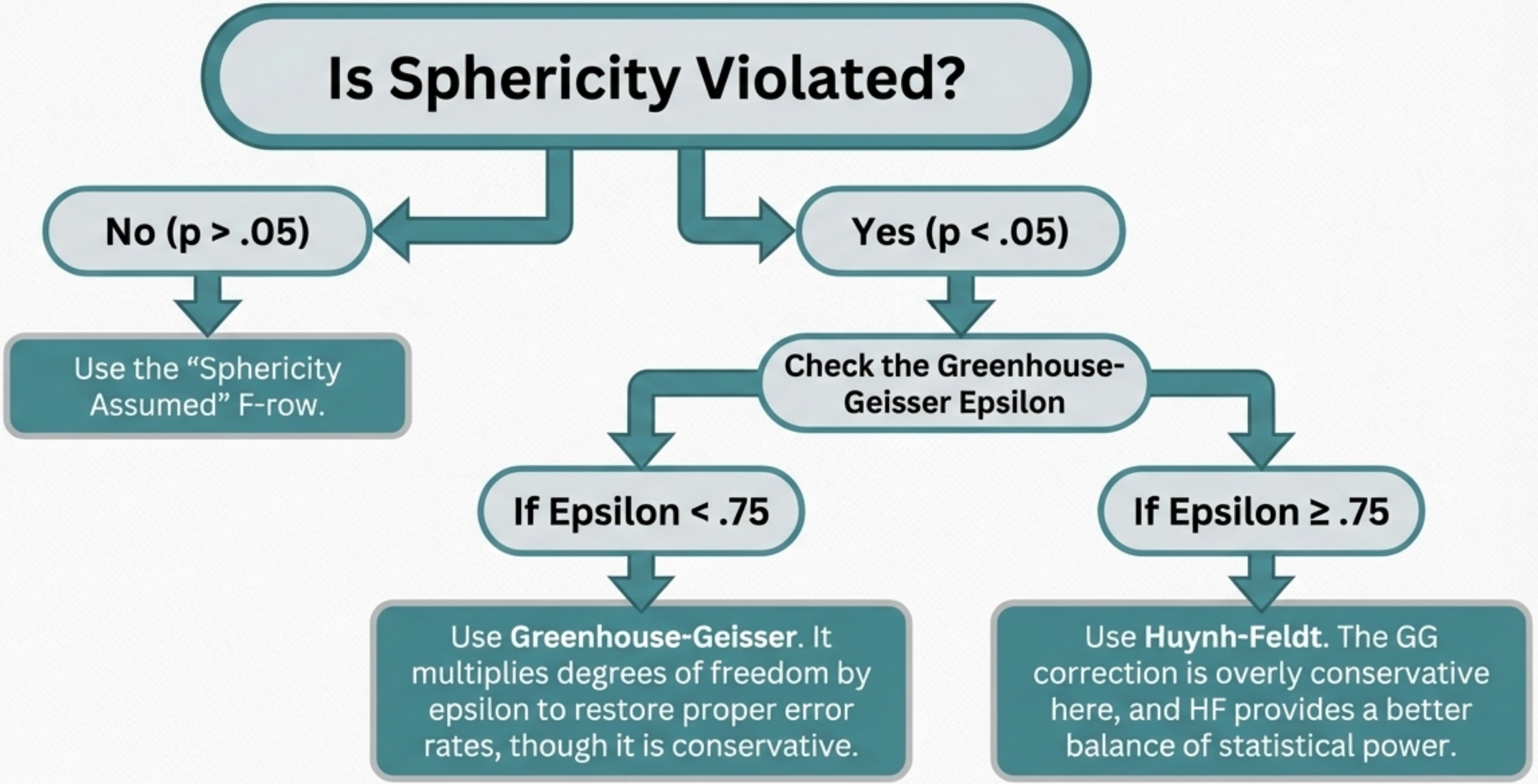
Mauchly's test evaluates whether the sphericity assumption holds. If $p > .05$, the assumption is met. If $p < .05$, the assumption is violated and a correction must be applied to prevent inflated Type I error.

Application

In our strength tracking data, Mauchly's test returned $W(2) = .932$ with $p = .054$. Because the p-value is greater than .05, we proceed with sphericity assumed.

NotebookLM

The Correction Matrix: Adjusting for Violations


Is Sphericity Violated?
No (p > .05)
Use the “Sphericity Assumed” F-row.
Yes (p < .05)
Check the Greenhouse-Geisser Epsilon
If Epsilon < .75
Use Greenhouse-Geisser. It multiplies degrees of freedom by epsilon to restore proper error rates, though it is conservative.
If Epsilon ≥ .75
Use Huynh-Feldt. The GG correction is overly conservative here, and HF provides a better balance of statistical power.

NotebookLM

Source ting Lab: **Interpreting the Omnibus F-Test**

Source	SS	df	MS	F	p
Time		2	––	116.0	< .001
Error (Time)		58	––		

The Translation

The omnibus test reveals a massive, statistically significant effect of time on muscular strength: $F(2, 58) = 116.0$, $p < .001$.

Application Note

A significant omnibus F only tells you that a meaningful change occurred somewhere across the 12 weeks. Post hoc tests are required to identify exactly which intervals drove the improvement.

NotebookLM

Locating the Change: Bonferroni-Corrected Post Hocs

To control the familywise error rate across multiple comparisons, the Bonferroni adjustment tightens the significance threshold.

Comparison	Mean Difference	p-value	Significance
Mid vs. Pre	+2.02 kg	p < .001	✓
Post vs. Mid	+3.36 kg	p < .001	✓
Post vs. Pre	+5.38 kg	p < .001	✓

Conclusion: Strength increased progressively and significantly at every measured stage of the 12-week program.

NotebookLM

Quantifying Magnitude: The Effect Size Matrix

$$\eta_p^2 = \frac{SS_{effect}}{SS_{effect} + SS_{error}}$$

Excludes between-subjects variance from the denominator. Because individual baseline noise is removed, this value is larger than full eta-squared.

Our data: Partial Eta-Squared = .80

$$\omega_p^2 = \frac{SS_{effect} - df_{effect} * MS_{error}}{SS_{total} + MS_{error}}$$

Corrects for the upward bias inherent in eta-squared, providing a more accurate population estimate for small samples.

Our data: Partial Omega-Squared ≈ .88

$$d = \frac{M_{diff}}{SD_{diff}}$$

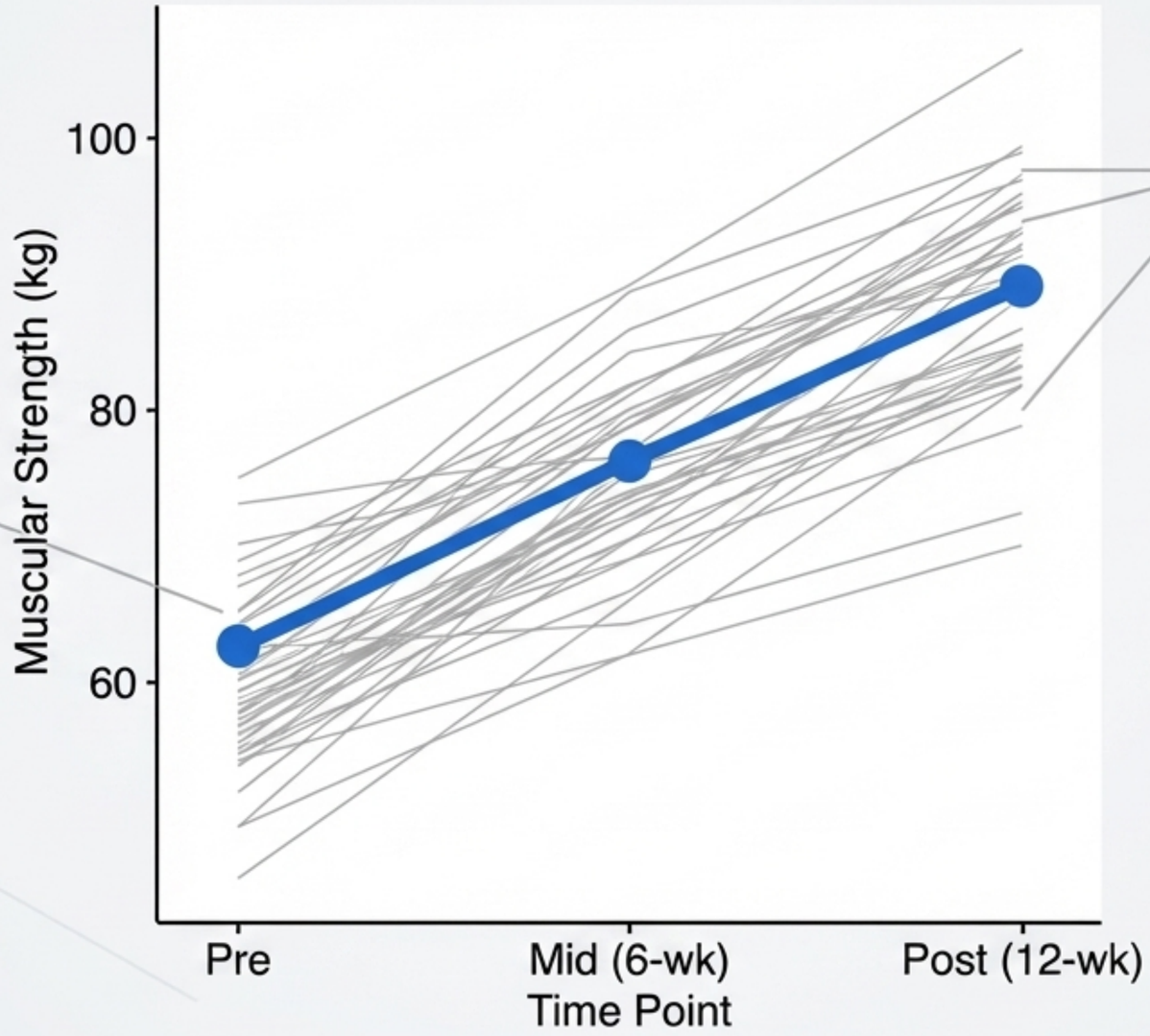
Computed specifically for the individual pairwise difference scores using the mean difference divided by the standard deviation of the differences.

Our data: Post vs Pre d = 2.97

NotebookLM

Visualizing Individual Trajectories

The Trend: The bold blue line represents the group mean showing progressive strength gains.



The Individual Lines: Each thin grey line is one participant. Because the lines move predominantly in the same upward direction without crisscrossing, the within-subject effect is highly consistent.

Synthesis Note:

Parallel trajectories equal small error variance. When participants respond uniformly to a treatment, statistical power surges.

NotebookLM

Synthesis: Assembling the APA Report

A one-way repeated measures ANOVA examined the effect of training time on muscular strength. Mauchly's test indicated that the sphericity assumption was not violated, [W(2) = .932, p = .054]. The within-subjects effect of time was statistically significant, [F(2, 58) = 116.0, p < .001], [η^2_p = .80]. Post hoc Bonferroni-corrected pairwise comparisons indicated that strength increased significantly from pre-training to mid-training...

Links to Slide 8: Sphericity Check

Links to Slide 10: Omnibus Test

Links to Slide 12: Effect Size

NotebookLM

Diagnostic Checklist: Common Pitfalls to Avoid

Never default to ‘Sphericity Assumed’

Routinely relying on this **F-row** without checking Mauchly’s test first will **inflate your Type I error rate.**

Do not run unapproved post hocs

Pairwise comparisons are only statistically justified if the omnibus **F-test** is significant first.

Qualify your effect sizes

Partial eta-squared is mathematically larger **than full eta-squared. Directly comparing the two misrepresents your findings.**

Do not confuse error terms

Repeated measures error reflects individual inconsistency across time, not total within-group variability.

NotebookLM